NORTH-HOLLAND

A Theorem of the Alternative for Multihomogeneous Functions and Its Relationship to Diagonal Scaling of Matrices*

Bahman Kalantari
Department of Computer Science
Rutgers University
New Brunswick, New Jersey 08903

Submitted by Hans Schneider

ABSTRACT

We prove that, given a multihomogeneous function satisfying some initial conditions, either it has a certain nonnegative zero over a given subspace, or an associated logarithmic barrier function has a constrained stationary point. Under convexity precisely one of these conditions is satisfied. The main ingredients of the proof are the derivation of significant properties of the constrained stationary points of the logarithmic barrier function, and their relationship to corresponding points of an associated Karmarkar potential function. Corollaries of the theorem include a duality for Karmarkar's canonical linear program, which happens to be a stronger version of Gordan's theorem; a more general duality than an existing one concerning the diagonal scaling of symmetric matrices, also shown to be a stronger version of Gordan's theorem; and a diagonal scalability result for a class of multihomogeneous polynomials which is more general than a previously known result on the scalability of positive multidimensional matrices. We also give an algorithmic proof of the theorem of the alternative through a projective algorithm which is a generalization of a modified Karmarkar linear programming algorithm, and in the context of nonnegative matrix scaling becomes a variant of the well-known RAS algorithm.

1. INTRODUCTION

Gordan's theorem of the alternative, proved in 1873 (see Dantzig [7]), is the following intuitively simple, yet profound result: Given a set of points in

*This research was supported in part by the National Science Foundation under grants CCR-88-07518 and CCR-9208371.

0024-3795/96/$15.00
SSDI 0024-3795(94)00162-7

$\mathbb{R}^m$, either the origin lies in the convex hull of these points, or there exists a hyperplane separating the origin from the given points. Algebraically, Gordan's theorem is as follows: Let B be an $m \times n$ matrix with real coefficients. Then $U(B) = \{x \in \mathbb{R}^n : Bx = 0,\ x \geqslant 0,\ x \neq 0\}$ is nonempty or $U'(B) = \{y \in \mathbb{R}^m : B^T y > 0\}$ is nonempty. It is easy to see that the two conditions are mutually exclusive. Surprisingly, Gordan's theorem can be used to establish its equivalence to the following nonintuitive statement (see Corollary 2.2): Let Q be an $n \times n$ real symmetric matrix satisfying $d_0^T Q d_0 > 0$ for some $d_0 \in \mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x > 0\}$. Then

$$V = \{x \in \mathbb{R}^n : \phi(x) = x^T Q x = 0,\ x \geqslant 0,\ x \neq 0\} \neq \varnothing \tag{a}$$

or

$$\exists \lambda, d \in \mathbb{R}^n_+ \text{ such that } DQDe = \lambda, \tag{b}$$

where $D = \operatorname{diag}(d) = \operatorname{diag}(d_1, \dots, d_n)$, and $e = (1, \dots, 1)^T \in \mathbb{R}^n$. If Q is positive semidefinite, then precisely one of these conditions is satisfied.

The problem of determining if V is nonempty is a fundamental problem in mathematical programming and theoretical computer science, since it is equivalent to linear programming for positive semidefinite Q, and is NP-complete for general Q with rational entries (see [11]). Given the duality relation described by "(a) or (b)," together with the fact that solving a general linear programming is equivalent to testing if (a) is satisfied, it is fair to say that Gordan's theorem is the first "interior-point" view into linear programming. Given this duality, what could have motivated an "interior-point algorithm" would have been the observation that given D in (b), the vector $\bar{d} = (1/\sqrt{2})De$ is a stationary point of the logarithmic barrier function

$$\psi(x) = \phi(x) - \sum_{i=1}^{n} \lambda_i \ln x_i.$$

One difficulty however would have been due to the fact that the vector λ in (b) is unknown. In fact it turns out that one has an infinite degree of freedom in selecting λ and that the following stronger duality holds: Condition (a) is satisfied or

$$\forall \lambda \in \mathbb{R}^n_+,\ \exists d_\lambda \in \mathbb{R}^n_+ \text{ such that } D_\lambda Q D_\lambda e = \lambda, \tag{b$'$}$$

where $D_\lambda = \operatorname{diag}(d_\lambda)$.

The above duality was essentially proved by Marshall and Olkin [23]. However, they were primarily interested in nonnegative matrix scaling, a problem that has been of interest at least since the thirties. Thus, the optimization of the logarithmic barrier function was not pursued. An interior-point method for linear programming which utilizes the above duality, while optimizing the logarithmic barrier function $\psi(x)$, is described in [19].

In 1984, Karmarkar [17] considered linear programming as the problem of testing if the following is satisfied:

$$V = \{x \in W : \phi(x) = c^T x = 0,\ x \geqslant 0,\ x \neq 0\} \neq \varnothing, \tag{a$'$}$$

where $W = \{x \in \mathbb{R}^n : Ax = 0\}$, A and $m \times n$ matrix of rank m satisfying $Ae = 0$, and it is assumed that $\phi(e) > 0$. In his projective algorithm Karmarkar makes use of the potential function

$$f(x) = \frac{\phi(x)}{\pi(x)},$$

where $\pi(x) = (\prod_{i=1}^n x_i)^{1/n}$. Karmarkar's algorithm essentially proves the following duality: Precisely one of the following two conditions holds true: not (a$'$), or

$$\forall d \in W \cap \mathbb{R}^n_+,\ \exists x_d \in W \cap \mathbb{R}^n_+ \text{ satisfying } e^T x_d = e^T d,\ f(x_d) \leqslant \gamma f(d), \tag{b$''$}$$

where γ is a number depending only on n and can be taken to be $(2\exp(-1))^{1/n}$ (see [10]). In Karmarkar's algorithm, given $d \in W \cap \mathbb{R}^n_+$ satisfying $c^T d > 0$, in $O(n^3)$ arithmetic operations one can either conclude that V is empty, or compute x_d satisfying the condition (b$''$). The algorithm consists of replacing d with x_d and repeating the above. Karmarkar's duality can be extended to the case where ϕ is any homogeneous function of degree K [i.e. $\phi(\alpha x) = \alpha^K \phi(x)$] which is also convex over W, where in the definition of the corresponding potential function the exponent $1/n$ gets replaced with K/n, and the corresponding γ becomes $\{[(K+1)/K]^K \exp(-1)\}^{1/n}$ (see [13]).

As it turns out, Karmarkar's V in (a$'$) is none other than a disguised form of Gordan's $U(B)$ (see [14]). Furthermore, Karmarkar's potential is closely related to the logarithmic barrier function

$$\psi(x) = c^T x - \sum_{i=1}^n \ln x_i,$$

and another duality can be stated for Karmarkar's V: Precisely one of the following two conditions holds: (a′), or

$$\forall\lambda \in \mathbb{R}^n_+,\ \exists d_\lambda \in W \cap \mathbb{R}^n_+ \text{ such that } P_{d_\lambda} \nabla\phi_{d_\lambda}(e) = P_{d_\lambda}\lambda, \qquad (\text{b}''')$$

where $P_{d_\lambda} = I - D_\lambda A^T(AD_\lambda^2A^T)^{-1}AD_\lambda$, $D_\lambda = \operatorname{diag}(d_\lambda)$, and $\phi_{d_\lambda}(x) = \phi(D_\lambda x)$. More generally, the duality "(a′) or (b‴)" still holds if the function $\phi(x) = c^Tx$ in (a′) is replaced with $\phi(x) = x^TQx$, in which case the corresponding duality will become a more general version of the duality "(a) or (b′)" (see Corollary 2.3).

This paper proves the duality "(a′) or (b‴); however, it does so in a much more general setting: Let $\phi(x) = \phi(x^{(1)}, \dots, x^{(m)})$, $m \geqslant 1$, be any continuously differentiable real-valued function defined over the nonnegative points of a subspace $W = W_1 \times \cdots \times W_m$ of $\mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$, and for each $j = 1, \dots, m$, homogeneous of positive degree K_j with respect to nonnegative points of W_j. We let V be the set of nonnegative $x = (x^{(1)}, \dots, x^{(m)}) \in W$, each of whose components $x^{(j)}$ is nontrivial, satisfying $\phi(x) = 0$. Assuming that $\phi(d_0) > 0$ for some positive $d_0 \in W$, and given any *admissible* vector $\lambda \in \mathbb{R}^n$ (see Definition 1.1), we prove $V \neq \varnothing$ or the logarithmic barrier function $\psi(x) = \phi(x) - \sum_{i=1}^n \lambda_i \ln x_i$ has a constrained stationary point in W. If ϕ is convex over W, precisely one of the two conditions holds true, and ψ has at most one constrained stationary point in W. Moreover, for any positive vector d in W, the constrained stationary points of ψ are preserved under the change of variable $x \leftarrow Dx$, where $D = \operatorname{diag}(d)$. In particular if $d \in W$, is a positive constrained stationary point of ψ corresponding to $\lambda = (K_1e^{(n_1)}, \dots, K_me^{(n_m)})$, where $e^{(k)} = (1, \dots, 1) \in \mathbb{R}^k$, then $P_d \nabla\phi_d(e^{(n)}) = \lambda$, where P_d is the orthogonal projection matrix with respect to the subspace $W_d = \{x : Dx \in W\}$, and $\phi_d(x) = \phi(Dx)$.

On one hand for $m = 1$, the problem of determining if V is nonempty is a fundamental problem in linear, quadratic, and more generally polynomial programming. Indeed, given any polynomial $p(x)$, $x \in \mathbb{R}^n$, and a polytope $P \subset \mathbb{R}^n$, the problem of checking if $p(x)$ attains a specific value can be reduced to the decision problem which asks if a homogeneous polynomial $\phi(x)$ has a nontrivial nonnegative zero in a subspace W. Such a transformation can be established in a similar fashion to the case of quadratics, described in [11]. On the other hand, for $m \geqslant 1$ the problem of determining if ψ has a desired stationary point is a fundamental problem in multihomogeneous programming which includes matrix and multidimensional matrix scaling. In addition to this theorem of the alternative, the paper describes a projective algorithm, which is a generalization of a modified Karmarkar algorithm, to find either a point in V or a desired constrained stationary point of ψ.

The proof of the theorem is based on the derivation of significant properties of the constrained stationary points of the logarithmic barrier function, and their intricate relationship to corresponding points of an associated Karmarkar potential function. Even for linear programming the duality "(a′) or (b‴)" gives a new and powerful duality. For instance, in [14], based on this duality, we describe a very simple polynomial-time linear programming algorithm. In the context of matrix scaling and when $\phi(x)$ is quadratic, the corresponding duality results in a more general version of Marshall and Olkin's duality. Furthermore, one can conclude a scalability result for a class of multihomogeneous polynomials which is more general than an existing result on the diagonal scalability of positive multidimensional matrices. The convergence proof of the modified projective algorithm establishes the true strength of the algorithm, whether or not ϕ is convex. For example, even in the restricted class of nonnegative multilinear forms, the projective algorithm becomes an alternative algorithm to the so-called RAS algorithm (e.g., see [30] for RAS).

In what follows we shall assume that ϕ is the multihomogeneous function described above. From multihomogeneity for any positive scalars α_j, $j = 1,\dots,m$, and any nonnegative $x = (x^{(1)},\dots,x^{(m)}) \in W = W_1 \times \cdots \times W_m$ we have

$$\phi\left(\alpha_1 x^{(1)},\dots,\alpha_m x^{(m)}\right) = \left(\prod_{j=1}^{m} \alpha_j^{K_j}\right)\phi(x). \tag{1.1}$$

We assume that W contains strictly positive points, and that if, for $j \in \{1,\dots,m\}$, W_j is a proper subspace of $\mathbb{R}^{n_j}$, then $W_j = \{x^{(j)} \in \mathbb{R}^{n_j} : A_j x^{(j)} = 0\}$, where A_j is a matrix of full row rank. The corresponding orthogonal projection matrix with respect to W_j is thus $P^{(j)} = I^{(j)} - A_j^T(A_j A_j^T)^{-1}A_j$, where $I^{(j)}$ is the $n_j \times n_j$ identity matrix. If $W_j = \mathbb{R}^{n_j}$, we can define $P^{(j)}$ to be $I^{(j)}$. Since $W = \{x \in \mathbb{R}^n : Ax = 0\}$, where $A = \operatorname{diag}(A_1,\dots,A_m)$, the orthogonal projection matrix P with respect to W satisfies

$$P = \operatorname{diag}(P^{(1)},\dots,P^{(m)}). \tag{1.2}$$

We first introduce the necessary notation that will be used throughout the paper.

NOTATION 1.1. For $x \in \mathbb{R}^n$ we write $x = (x^{(1)},\dots,x^{(m)})$ to mean that $x^{(j)} \in \mathbb{R}^{n_j}$, $j = 1,\dots,m$. The partial gradient of $\phi(x)$ with respect to $x^{(j)}$

will be denoted by $\nabla_j \phi$. Given a positive vector $d = (d^{(1)}, \ldots, d^{(m)}) \in W$, for $j = 1, \ldots, m$ we define the diagonal matrix $D_j = \operatorname{diag}(d^{(j)}) = \operatorname{diag}(d_1^{(j)}, \ldots, d_{n_j}^{(j)})$. Thus, $D = \operatorname{diag}(d) = \operatorname{diag}(D_1, \ldots, D_m)$. The change of variable $x \leftarrow Dx$, induces a new multihomogeneous function $\phi_d(x) = \phi(Dx)$ satisfying (1.1), and a new subspace $W_d = \{x \in \mathbb{R}^n : ADx = 0\}$, whose orthogonal projection matrix as in (1.2) is given by $P_d = \operatorname{diag}(P_d^{(1)}, \ldots, P_d^{(m)})$, where for $j = 1, \ldots, m$, $P_d^{(j)} = I - D_j A_j^T (A_j D_j^2 A_j^T)^{-1} A_j D_j$. For a given k we write $\mathbb{R}_+^k = \{x \in \mathbb{R}^k : x > 0\}$. Given $x, \lambda \in \mathbb{R}_+^k$, we define $x^{-1} = (1/x_1, \ldots, 1/x_k)^T$, and $\lambda \circ x^{-1} = (\lambda_1/x_1, \ldots, \lambda_k/x_k)^T$. The vector of ones in $\mathbb{R}^k$ is denoted by $e^{(k)}$. The Euclidean l_1 and l_2 norms will be denoted by $\|\cdot\|_1$ and $\|\cdot\|$, respectively.

The multihomogeneous function ϕ satisfies the following two important properties:

PROPERTY 1.1 (Multihomogeneity of partial gradients). *For any set of positive scalars α_j, $j = 1, \ldots, m$, and $x = (x^{(1)}, \ldots, x^{(m)}) \in W \cap \mathbb{R}_+^n$,*

$$\nabla_j \phi\left(\alpha_1 x^{(1)}, \ldots, \alpha_m x^{(m)}\right) = \frac{1}{\alpha_j} \left(\prod_{i=1}^{m} \alpha_i^{K_i} \right) \nabla_j \phi(x).$$

PROPERTY 1.2 (Euler's equation). *For any nonnegative $x = (x^{(1)}, \ldots, x^{(m)}) \in W$,*

$$x^{(j)T} \nabla_j \phi(x) = K_j \phi(x) \qquad \forall j = 1, \ldots, m.$$

For a given $\lambda \in \mathbb{R}_+^n$, consider the logarithmic barrier function

$$\psi(x) = \phi(x) - \sum_{i=1}^{n} \lambda_i \ln x_i.$$

Before presenting the precise statement of the main theorem, we shall first arrive at two simple but important properties regarding the constrained stationary points of ψ. The first of these properties (Lemma 1.1) pertains to the multihomogeneity of ϕ, while the second (Lemma 1.2) is solely a property of the logarithmic barrier function $B(x) = -\sum_{i=1}^{n} \ln x_i$.

Suppose that $d_\lambda = (d_\lambda^{(1)}, \dots, d_\lambda^{(m)}) \in W \cap \mathbb{R}_+^n$ is a constrained stationary point of ψ, i.e.

$$\nabla\psi(d_\lambda) = \nabla\phi(d_\lambda) - \lambda \circ d_\lambda^{-1} = A^T v, \tag{1.3}$$

where v is the vector of Lagrange multipliers. From (1.3), v can be computed to be

$$v = (AA^T)^{-1} A\left[\nabla\phi(d_\lambda) - \lambda \circ d_\lambda^{-1}\right]. \tag{1.4}$$

Equivalently, from (1.3) and (1.4), we get

$$P\,\nabla\phi(d_\lambda) = P\left(\lambda \circ d_\lambda^{-1}\right). \tag{1.5}$$

From (1.2) we conclude that (1.5) is equivalent to

$$P^{(j)}\nabla_j\phi(d_\lambda) = P^{(j)}\left(\lambda^{(j)} \circ d_\lambda^{(j)^{-1}}\right) \qquad \forall j = 1, \dots, m. \tag{1.6}$$

Multiplying both sides of (1.6) by $d_\lambda^{(j)\,T}$, using the fact that $P^{(j)}d_\lambda^{(j)} = d_\lambda^{(j)}$, and applying Property 1.2, we get

$$K_j\phi(d_\lambda) = \sum_{i=1}^{n_j} \lambda_i^{(j)} = \|\lambda^{(j)}\|_1 \qquad \forall j = 1, \dots, m. \tag{1.7}$$

This brings us to the following definition.

DEFINITION 1.1. We shall say that $\lambda = (\lambda^{(1)}, \dots, \lambda^{(m)}) \in \mathbb{R}^n$ is *admissible* if $\lambda > 0$ and if the ratios $K_j/\|\lambda^{(j)}\|_1$, $j = 1, \dots, m$, are identical.

From (1.3)–(1.7) and Definition 1.1, we thus have

LEMMA 1.1. *Let $\lambda \in \mathbb{R}_+^n$ be a given vector. Suppose that $\psi(x) = \phi(x) - \sum_{i=1}^n \lambda_i \ln x_i$ has a constrained stationary point d_λ in $W \cap \mathbb{R}_+^n$. Then λ is admissible. Moreover, if σ is the common value of $K_j/\|\lambda^{(j)}\|_1$, then $\phi(d_\lambda) = \sigma^{-1}$.*

Suppose (1.3) holds. Let d be any given point in $W \cap \mathbb{R}^n_+$. Multiplying (1.3) by $D = \mathrm{diag}(d)$, observing that $D(\lambda \circ d_\lambda^{-1}) = \lambda \circ (D^{-1}d_\lambda)^{-1}$, and noting that from the chain rule $D\,\nabla\phi(d_\lambda) = \nabla\phi_d(D^{-1}d_\lambda)$, we get

$$\nabla\phi_d\big(D^{-1}d_\lambda\big) - \lambda \circ \big(D^{-1}d_\lambda\big)^{-1} = DA^T v, \tag{1.8}$$

implying that the point $d'_\lambda = D^{-1}d_\lambda$, which lies in $W_d \cap \mathbb{R}^n_+$, is a constrained stationary point of

$$\psi_d(x) = \psi(Dx) = \phi(Dx) - \sum_{i=1}^n \lambda_i \ln x_i d_i$$

$$= \phi_d(x) - \sum_{i=1}^n \lambda_i \ln x_i - \sum_{i=1}^n \lambda_i \ln d_i. \tag{1.9}$$

Clearly, the converse also holds. Furthermore, by solving for v in (1.8) we can obtain as equation analogous to (1.5). More precisely, we have

LEMMA 1.2. *Let $\lambda \in \mathbb{R}^n$ be a given admissible vector, and $\psi(x) = \phi(x) - \sum_{i=1}^n \lambda_i \ln x_i$. A point $d_\lambda \in W \cap \mathbb{R}^n_+$ is a constrained stationary point of $\psi(x)$ if and only if for any d in $W \cap \mathbb{R}^n_+$, the point $d'_\lambda = D^{-1}d_\lambda \in W_d \cap \mathbb{R}^n_+$ is a constrained stationary point of $\psi_d(x)$ if and only if*

$$P_d\, \nabla\phi_d(d'_\lambda) = P_d\big(\lambda \circ d'^{-1}_\lambda\big). \tag{I}$$

In particular, if for $\lambda = (K_1 e^{(n_1)}, \ldots, K_m e^{(n_m)})$, d_λ is a constrained stationary point of $\psi(x)$, then

$$P_{d_\lambda}\, \nabla\phi_{d_\lambda}(e) = \lambda, \tag{II}$$

where $e = (e^{(n_1)}, \ldots, e^{(n_m)})$, the vector of ones in $\mathbb{R}^n$.

Having stated the two properties of $\psi(x)$, we now state the main theorem.

THEOREM 1.1. *Assume* $\phi(d_0) > 0$ *for some* $d_0 \in W \cap \mathbb{R}^n_+$. *Let* $\lambda \in \mathbb{R}^n$ *be a given admissible vector, and* $\psi(x) = \phi(x) - \sum_{i=1}^n \lambda_i \ln x_i$. *Then either*

$$V = \{x \in W : \phi(x) = 0,\ x^{(j)} \geqslant 0,\ x^{(j)} \neq 0,\ j = 1, \dots, m\} \neq \varnothing \quad \text{(i)}$$

or

$$\psi(x) \text{ has a constrained stationary point } d_\lambda \in W \cap \mathbb{R}^n_+. \quad \text{(ii)}$$

If ϕ *is convex over* W, *then (i) and (ii) are exclusive, and* d_λ, *if it exists, is unique.*

REMARK 1.1. In Lemma 1.2 the proof of the first "if and only if" is merely the application of the chain rule and holds for any differentiable function ψ. However, the correctness of the second "if and only if" part of the lemma can be viewed as a remarkable property of the logarithmic barrier function $B(x) = -\sum_{i=1}^n \ln x_i$ and it would no longer be true if we had selected another barrier function, say $\sum_{i=1}^n 1/x_i$. This distinguishing property of the logarithmic barrier function (which is also used in Karmarkar's algorithm, but in a transparent fashion) and the multihomogeneity of $\phi(x)$ are the two most important ingredients in the proof of Theorem 1.1, as well as the proof of convergence properties of the projective algorithm.

Equation (I) of Lemma 1.2 can be viewed as an algebraic description of condition (ii) of Theorem 1.1. The algebraic description is useful in several ways. Firstly, we shall use it to prove the exclusivity of conditions (i) and (ii) of the theorem. Secondly, the statements of the corollaries of the theorem are all with respect to this algebraic description. It is interesting that in the vast literature of nonnegative matrix scaling, in which $W = \mathbb{R}^n$, the equivalence of $DQDe = \lambda$ to $\bar{d} = (1/\sqrt{2})De$ being a stationary point of $\psi(x) = x^TQx - \sum_{i=1}^n \lambda_i \ln x_i$ was not noticed or made use of until the work of Marshall and Olkin [23].

In Section 2, we give some immediate corollaries of Theorem 1.1. In Section 3, we give the proof of that theorem. In Section 4, we give a projective algorithm that finds either a point of V or a constrained stationary point of ψ in $W \cap \mathbb{R}^n_+$.

2. SOME COROLLARIES OF THE MAIN THEOREM

In this section we describe some immediate and important corollaries of Theorem 1.1.

COROLLARY 2.1. *Let ϕ, d_0, and V be as in Theorem 1.1. Then either $V \neq \emptyset$ or there exists $d \in W \cap \mathbb{R}^n_+$ satisfying $P_d \nabla\phi_d(e) > 0$. Under the assumption of convexity, precisely one of these conditions is true.*

Proof. If $V = \emptyset$, then from Theorem 1.1 and Lemma 1.2, for $\lambda = (K_1 e^{(n_1)}, \ldots, K_m e^{(n_m)})$ Equation (II) of Lemma 1.2 holds. In particular $P_{d_\lambda} \nabla\phi_{d_\lambda}(e) > 0$. Conversely, suppose there exists $d \in W \cap \mathbb{R}^n_+$ satisfying $P_d \nabla\phi_d(e) > 0$. Let $\lambda = P_d \nabla\phi_d(e)$. Since $P_d^2 = P_d$, we have $P_d\lambda = \lambda$. Also, $e \in W_d$. Thus, for this λ Equation (I) of Lemma 1.2 holds for $d'_\lambda = e$, and $d_\lambda = d$ is a constrained stationary point of the corresponding $\psi(x)$. The admissibility of λ follows from Lemma 1.1. ■

REMARK 2.1. Corollary 2.1 can be viewed as a generalization of Gordan's theorem. If $W = \mathbb{R}^n$, then $P = I$ and from the chain rule the condition $P_d \nabla\phi_d(e) > 0$ can be replaced with $\nabla\phi(d) > 0$. In particular, for $m = 1$ and $\phi(x) = x^TQx$, the corollary becomes equivalent to Gordan's theorem (Corollary 2.2). For the case where W is a proper subspace the condition $P_d \nabla\phi_d(e) > 0$ cannot be replaced with $P\nabla\phi(d) > 0$. This can be checked for instance for the case where $m = 1$, $\phi(x) = c^Tx$, $c = (4, 1, 1)^T$, and $A = [1, -1, 1]$. In the case of $m = 1$, $\phi(x) = c^Tx$, by making use of the corresponding duality, which happens to be a stronger version of Gordan's theorem, a simple polynomial-time linear programming algorithm is described in [14].

COROLLARY 2.2 (Equivalent of Gordan's theorem). *Let $\phi(x) = x^TQx$, where Q is a symmetric $n \times n$ matrix satisfying $d_0^TQd_0 > 0$ for some $d_0 > 0$, and $V = \{x \in \mathbb{R}^n : \phi(x) = 0,\ x \geqslant 0,\ x \neq 0\}$. Then either $V \neq \emptyset$ or there exist $d, \lambda \in \mathbb{R}^n_+$ satisfying $DQDe = \lambda$. If Q is positive semidefinite, precisely one of these conditions is satisfied.*

Proof of equivalence of Gordan's theorem. By taking $Q = B^TB$, where B is the matrix of Gordan's theorem, it is easy to check that Corollary 2.2 implies Gordan's theorem. Now we show that Corollary 2.2 can also be directly deduced from Gordan's theorem. Assume that $V = \emptyset$. Then it follows that $Qx + y = 0$ has no nonnegative nontrivial solution (x, y). Otherwise, x would have to be nontrivial and $x^TQx = -x^Ty \leqslant 0$. Since $\phi(x) = 0$ implies $V \neq \emptyset$, a contradiction, we must have $\phi(x) < 0$. But the assumption that $\phi(d_0) > 0$ implies that for some convex combination of x and d_0, ϕ is zero, again a contradiction. Now, applying Gordan's theorem to the matrix $B = [Q, I]$, I the identity matrix, we obtain the satisfiability of the system $Qd > 0$, $d > 0$. Multiplying $Qd > 0$ by $D = \operatorname{diag}(d)$, we get the result. The fact that under positive semidefiniteness of Q both conditions cannot occur is trivial, since $\phi(x) = 0$ if and only if $Qx = 0$. ■

A stronger version of Corollary 2.2 is the following corollary of Theorem 1.1.

COROLLARY 2.3. *Let $\phi(x) = x^T Qx$, where Q is a symmetric $n \times n$ matrix satisfying $d_0^T Q d_0 > 0$ for some $d_0 \in W \cap \mathbb{R}^n_+$, where $W = \{x \in \mathbb{R}^n : Ax = 0\}$, A an $m \times n$ matrix of rank m. Then either $V = \{x \in W : \phi(x) = 0,\ x \geqslant 0,\ x \neq 0\} \neq \varnothing$, or, given $\lambda \in \mathbb{R}^n_+$, there exists $d_\lambda \in W \cap \mathbb{R}^n_+$ satisfying $P_{d_\lambda} D_\lambda Q D_\lambda P_{d_\lambda} e = P_{d_\lambda} \lambda$. If Q is positive semidefinite, precisely one of these conditions is satisfied. In particular, if $V = \varnothing$ and $\lambda = e$, the matrix $P_{d_\lambda} D_\lambda Q D_\lambda P_{d_\lambda}$ is doubly quasistochastic.* ■

REMARK 2.2. In Corollary 2.2, to determine if V is nonempty is NP-complete; see Kalantari [11]. For Q positive semidefinite the problem is equivalent to linear programming and hence polynomially solvable for rational inputs; e.g., see Khachiyan [18], Karmarkar [17], Kalantari [11], and Khachiyan and Kalantari [19]. The problem of testing for the existence of $\lambda > 0$ such that for some $d > 0$ one has $DQDe = \lambda$ is solvable in polynomial time, since it is equivalent to a feasibility problem in linear programming. Given a fixed $\lambda > 0$, the problem of testing the solvability of $P_{d_\lambda} D_\lambda Q D_\lambda P_{d_\lambda} e = P_{d_\lambda} \lambda$, for some $d_\lambda \in W \cap \mathbb{R}^n_+$, is polynomially solvable if Q is positive semidefinite, and is NP-hard in general. For $W = \mathbb{R}^n$, $\lambda = e$, these are proved in Khachiyan and Kalantari [19] and Khachiyan [20], respectively. For the special case of $W = \mathbb{R}^n$, Corollary 2.3 was essentially first proved by Marshall and Olkin [23]. For general W and $\lambda = e$, Corollary 2.3 was proved in [11].

The next corollary of Theorem 1.1 is a more general version of Corollary 2.3 stated for two-dimensional and multidimensional matrices and for $W = \mathbb{R}^n$.

COROLLARY 2.4. *Let $Q = (q_{i_1 \dots i_m})$, $i_1, \dots, i_m = 1, \dots, N$, be an m-dimensional matrix with entries $q_{i_1 \dots i_m}$. Let $\lambda^{(1)}, \dots, \lambda^{(m)} \in \mathbb{R}^N$ be any set of positive vectors with a common l_1 norm. For $x = (x^{(1)}, \dots, x^{(m)})$ where $x^{(j)} \in \mathbb{R}^N$, consider the multilinear form $\phi(x) = \sum_{i_1=1}^N \cdots \sum_{i_m=1}^N q_{i_1 \dots i_m} x_{i_1}^{(1)} \cdots x_{i_m}^{(m)}$. If $V = \{x : x^{(j)} \in \mathbb{R}^N,\ \phi(x) = 0,\ x^{(j)} \geqslant 0,\ x^{(j)} \neq 0,\ j = 1, \dots, m\}$ is empty, then there exist positive vectors $d^{(1)}, \dots, d^{(m)} \in \mathbb{R}^N$ such that for* *all* $j = 1, \ldots, m$ *and* $i_j = 1, \ldots, N$, *the following* scaling *equations are satisfied*:

$$\sum_{i_1=1}^{N} \cdots \sum_{i_{j-1}=1}^{N} \sum_{i_{j+1}=1}^{N} \cdots \sum_{i_m=1}^{N} q_{i_1 \cdots i_m} d_{i_1}^{(1)} \cdots d_{i_m}^{(m)} = \lambda_{i_j}^{(j)},$$

in which case Q *is said to be* scalable. *In particular, if all entries of* Q *are positive, then* Q *is scalable.*

Proof. Observe that the above scaling equations are equivalent to the equations $\nabla_{i_j} \psi(d_\lambda) = 0$ for the corresponding ψ of Theorem 1.1. Suppose all entries of Q are positive. In this case, if for a given $x \geqslant 0$ we have $x^{(j)} \neq 0$ for all $j = 1, \ldots, m$, then for some set of indices $i_1, \ldots, i_m \in \{1, \ldots, N\}$ we have $x_{i_1}^{(1)} \cdots x_{i_m}^{(m)} \neq 0$. This implies that V is empty; hence from Theorem 1.1, Q is scalable. ■

REMARK 2.3. The problem of scaling of positive and more generally nonnegative matrices is an extremely rich and applied problem which has been studied extensively; e.g. see [1, 5, 6, 8, 21–24, 26–33]. The polynomial-time solvability of the problem has been recently shown in [15]. In particular, the scalability of nonnegative matrices based on the structural pattern of nonzero entries is well known; e.g. see [5, 8, 21, 26, 28, 33]. These structural properties have also been generalized to nonnegative multidimensional matrices (e.g. see [3, 9]): Let $Q = (q_{i_1 \ldots i_m})$ be any m-dimensional matrix with nonnegative entries. An m-dimensional nonnegative matrix $B = (b_{i_1 \ldots i_m})$ is said to have the same *pattern* as Q if an entry of B is zero if and only if the corresponding entry of Q is zero. Then Q is scalable if and only if there exists B with the same pattern as Q such that for all $j = 1, \ldots, m$ and $i_j = 1, \ldots, N$ the following equations are satisfied:

$$\sum_{i_1=1}^{N} \cdots \sum_{i_{j-1}=1}^{N} \sum_{i_{j+1}=1}^{N} \cdots \sum_{i_m=1}^{N} b_{i_1 \ldots i_m} = \lambda_{i_j}^{(j)}.$$

While for positive multidimensional matrices Corollary 2.4 is well known, in the presence of a subspace W (as in the case of Corollary 2.3) the corresponding scalability equations give a new result. Furthermore, one can state a more general version of Corollary 2.4 for the case there $\phi(x)$ in Theorem 1.1 is a multihomogeneous polynomial and, with an analogous proof to that of Corollary 2.4, conclude a scalability result for the subclass of such functions that can be termed "positive multihomogeneous polynomials," i.e. functions $\phi(x)$ having a term with a positive coefficient corresponding to any set of indices $i_1 \in \{1, \dots, n_1\}, \dots, i_m \in \{1, \dots, n_m\}$.

3. PROOF OF THEOREM 1.1

Let $\lambda = (\lambda^{(1)}, \dots, \lambda^{(m)}) \in \mathbb{R}^n$ be a given admissible vector. Denote the common ratio $K_j / \|\lambda^{(j)}\|_1$ by σ. For $x = (x^{(1)}, \dots, x^{(m)}) \in W \cap \mathbb{R}^n_+$, define

$$\pi(x) = \prod_{j=1}^{m} \pi_j(x^{(j)}),$$

where

$$\pi_j(x^{(j)}) = \left(\prod_{i_j=1}^{n_j} \left(x_{i_j}^{(j)} \right)^{\lambda_{i_j}^{(j)}} \right)^{\sigma} \qquad \forall j = 1, \dots, m.$$

Define the (Karmarkar) potential function

$$f(x) = \frac{\phi(x)}{\pi(x)}.$$

Note that $\pi_j(x^{(j)})$ is homogeneous of degree K_j over $\mathbb{R}^{n_j}$, $j = 1, \dots, m$. Thus, f is multihomogeneous of degree zero, i.e., for any positive scalars α_j, $j = 1, \dots, m$,

$$f\left(\alpha_1 x^{(1)}, \dots, \alpha_m x^{(m)}\right) = f(x). \tag{3.1}$$

We first prove that positive constrained stationary points of ψ and f are within scalar multiples of each other.

LEMMA 3.1. *Suppose $P\,\nabla\psi(\hat{x}) = 0$, $\hat{x} \in W \cap \mathbb{R}^n_+$. Then $P\,\nabla f(\alpha\hat{x}) = 0$ for all positive scalars α. Conversely, suppose $P\,\nabla f(\hat{x}) = 0$ for some $\hat{x} \in W \cap \mathbb{R}^n_+$ satisfying $\phi(\hat{x}) > 0$. Then there exists $\alpha_* > 0$ such that $P\,\nabla\psi(\alpha_*\hat{x}) = 0$.*

Proof. Let $K = \sum_{j=1}^{m} K_j$. Recalling Notation 1.1, for any $x = (x^{(1)}, \dots, x^{(m)}) \in W \cap \mathbb{R}_+^n$ we have

$$\nabla_j \pi_j(x^{(j)}) = \sigma \pi_j(x^{(j)})(\lambda^{(j)} \circ x^{(j)^{-1}}) \qquad \forall j = 1, \dots, m,$$

from which it follows that

$$\nabla_j f(x) = \frac{1}{\pi(x)} \left[\nabla_j \phi(x) - \sigma \phi(x)(\lambda^{(j)} \circ x^{(j)^{-1}}) \right] \qquad \forall j = 1, \dots, m. \tag{3.2}$$

Suppose $P\, \nabla \psi(\hat{x}) = 0$ for some $\hat{x} \in W \cap \mathbb{R}_+^n$. Equivalently, from (1.2), $P^{(j)}\, \nabla_j \psi(\hat{x}) = 0$ for all $j = 1, \dots, m$. Also, from Lemma 1.1 we have $\phi(\hat{x}) = \sigma^{-1}$. From this and (3.2) we get

$$\nabla_j f(\hat{x}) = \frac{1}{\pi(\hat{x})} \nabla_j \psi(\hat{x}) \qquad \forall j = 1, \dots, m.$$

Thus, $P^{(j)}\, \nabla_j f(\hat{x}) = 0$ for all $j = 1, \dots, m$. It is easy to check that for $x > 0$ and any scalar $\alpha > 0$ we have $\nabla_j f(\alpha x) = (1/\alpha) \nabla_j f(x)$ (indeed, Property 1.1 also applies to f), so that $P^{(j)}\, \nabla_j f(\alpha \hat{x}) = (1/\alpha) P^{(j)}\, \nabla_j f(\hat{x}) = 0$ for all $j = 1, \dots, m$. Equivalently, $P\, \nabla f(\alpha \hat{x}) = 0$.

Now suppose $P\, \nabla f(\hat{x}) = 0$ for some $\hat{x} \in W \cap \mathbb{R}_+^n$ satisfying $\phi(\hat{x}) > 0$. Equivalently

$$P^{(j)}\, \nabla_j f(\hat{x}) = 0 \qquad \forall j = 1, \dots, m. \tag{3.3}$$

Letting $\delta = \sigma \phi(\hat{x})$, from (3.2) and (3.3) we get

$$P^{(j)} \left[\nabla_j \phi(\hat{x}) - \delta(\lambda^{(j)} \circ \hat{x}^{(j)^{-1}}) \right] = 0 \qquad \forall j = 1, \dots, m. \tag{3.4}$$

For any $\alpha > 0$, multiplying (3.4) by α^{K-1} and using Property 1.1, we get

$$P^{(j)} \left\{ \nabla_j \phi(\alpha \hat{x}) - \alpha^K \delta \left[\lambda^{(j)} \circ (\alpha \hat{x}^{(j)})^{-1} \right] \right\} = 0 \qquad \forall j = 1, \dots, m.$$

Since $\delta > 0$, the equation $\alpha^K \delta = 1$ has a positive solution α_*, and from the above we get $P^{(j)} \nabla_j \psi(\alpha_* \hat{x}) = 0$ for all $j = 1, \ldots, m$. Equivalently, $P \nabla \psi(\alpha_* \hat{x}) = 0$. ■

Define

$$S = \left\{ x \in \mathbb{R}^n : e^{(n_j)^T} x^{(j)} = n_j,\ x^{(j)} \geqslant 0,\ j = 1, \ldots, m \right\},$$

$$S_+ = \{ x \in S : x > 0 \},$$

$$\phi^* = \min\{ \phi(x) : x \in W \cap S \}, \tag{3.5}$$

$$f^* = \inf\{ f(x) : x \in W \cap S_+ \}.$$

LEMMA 3.2. *$V \neq \varnothing$ if and only if $\phi^* \leqslant 0$.*

Proof. Suppose $x \in V$. For $j = 1, \ldots, m$, define $\alpha_j = n_j / e^{(n_j)^T} x^{(j)}$. Then $(\alpha_1 x^{(1)}, \ldots, \alpha_m x^{(m)})$ is in $W \cap S$, and from (1.1), $\phi(\alpha_1 x^{(1)}, \ldots, \alpha_m x^{(m)}) = 0$. Thus, $\phi^* \leqslant 0$. The converse follows from the continuity of ϕ, the convexity of $W \cap S$, and the assumption that $\phi(d_0) > 0$. ■

LEMMA 3.3. *If $V = \varnothing$, then there exists $d_\lambda^* \in W \cap S_+$ such that $f^* = f(d_\lambda^*) > 0$.*

Proof. From Lemma 3.2, $V = \varnothing$ implies $\phi^* > 0$. Since $K_j > 0$ for all $j = 1, \ldots, m$, $f(x)$ approaches infinity as x approaches a boundary point of $W \cap S$. This implies that f^* is attained at a point $d_\lambda^* \in W \cap S_+$. ■

LEMMA 3.4. *If d is a local minimum of f over $W \cap S_+$, then d is a local minimum of f over $d \in W \cap \mathbb{R}_+^n$.*

Proof. Let $d_k = (d_k^{(1)}, \ldots, d_k^{(m)})$, $k = 1, 2, \ldots$, be a sequence of points in $W \cap \mathbb{R}_+^n$ converging to d. Then the sequence

$$\hat{d}_k = \left(\frac{n_1 d_k^{(1)}}{e^{(n_1)^T} d_k^{(1)}}, \ldots, \frac{n_m d_k^{(m)}}{e^{(n_m)^T} d_k^{(m)}} \right), \qquad k = 1, 2, \ldots,$$

is a sequence of points in $W \cap S_+$ which also converges to d. Moreover, from (3.1) we have $f(\hat{d}_k) = f(d_k)$, for all k. ■

From the application of Lemma 3.4 to d_λ^* (defined in Lemma 3.3), first-order optimality conditions, and Lemma 3.1, there exists $\alpha_* > 0$, such that if $d_\lambda = \alpha_* d_\lambda^*$, then $P\,\nabla\psi(d_\lambda) = 0$. This completes the proof of "(i) or (ii)" of Theorem 1.1.

Next we prove the exclusivity of (i) and (ii) under the assumption of convexity of $\phi(x)$. To do this we use the equivalence of (ii) to Equation (I) of Lemma 1.2. We have already shown that if (i) is not satisfied (i.e. V is empty), then (ii) is satisfied. Thus, we only need to show that if (ii) is true, then V is empty. So assume that (ii) is satisfied. By setting $d = d_\lambda$ in Equation (I) of Lemma 1.2 we get

$$P_{d_\lambda} \nabla\phi_{d_\lambda}(e) = P_{d_\lambda}\lambda. \tag{3.6}$$

Define

$$\phi_{d_\lambda}^* = \min\{\phi_{d_\lambda}(x) : x \in W_{d_\lambda} \cap S_\lambda\}, \tag{3.7}$$

where

$$S_\lambda = \{x \in \mathbb{R}^n : \lambda^{(j)T}x^{(j)} = \lambda^{(j)T}e^{(n_j)},\ x^{(j)} \geqslant 0,\ j = 1,\dots,m\}.$$

Note that (3.6) is precisely the first-order optimality condition at e with respect to the optimization problem defined in (3.7). Since $\phi(x)$ is convex over W, $\phi_{d_\lambda}(x)$ is convex over W_{d_λ}. Thus, (3.7) is a convex program, and (3.6) is the global optimality condition at e. Thus, $\phi_{d_\lambda}^* = \phi_{d_\lambda}(e)$. Multiplying (3.6) by e^T and applying Property 1.2 to $\phi_{d_\lambda}(x)$, which is multihomogeneous over the nonnegative points of W_{d_λ}, gives $\phi_{d_\lambda}(e) > 0$. Analogously to the proof of Lemma 3.1, it can easily be shown that $\phi_{d_\lambda}^* > 0$ implies the set $V_{d_\lambda} = \{x \in W_{d_\lambda} \cap \mathbb{R}^n : \phi_{d_\lambda}(x) = 0,\ x^{(j)} \geqslant 0,\ x^{(j)} \neq 0,\ j = 1,\dots,m\}$ is empty. But clearly, $V = \varnothing$ if and only if $V_{d_\lambda} = \varnothing$. Hence, the proof of exclusivity of (i) and (ii) under the assumption of the convexity of ϕ follows.

Finally, the uniqueness of d_λ under convexity of ϕ is a consequence of strict convexity of $B(x) = -\sum_{i=1}^n \ln x_i$. This completes the proof of Theorem 1.1.

4. A PROJECTIVE ALGORITHM

In this section we give a projective algorithm which finds either a point in V or a positive constrained stationary point of f (equivalently, by Lemma 3.1, a scalar multiple of a positive constrained stationary point of ψ). The algorithm can be thought of as a steepest-descent algorithm that uses projective transformations to center the current iterate.

Recalling the definition of S [see (3.5)], for a given $d \in W \cap S_+$ satisfying $f(d) > 0$, the iterative step of the algorithm performs a line search for $\phi_d(x)$ and the function

$$f_d(x) = \frac{\phi_d(x)}{\pi(x)}$$

in the direction of the steepest descent of f_d at $e = (1, \dots, 1)^T \in \mathbb{R}^n$. It updates d and repeats the above if necessary. The algorithm either finds a point $d \in W \cap S$ for which the set

$$V_d = \left\{x \in W_d \cap \mathbb{R}^n : \phi_d(x) = 0,\ x^{(j)} \geqslant 0,\ x^{(j)} \neq 0,\ j = 1, \dots, m\right\}$$

is nonempty, or obtains a point $d \in W \cap S_+$ satisfying $P_d \nabla f_d(e) = 0$, $\phi_d(e) > 0$. In the former case the algorithm terminates with a point in V (since $V_d \neq \emptyset$ if and only if $V \neq \emptyset$). In the latter case, we invoke the following lemma:

LEMMA 4.1. *Let d be a point in $W \cap \mathbb{R}^n_+$ with $\phi(d) > 0$. Then $P_d \nabla f_d(e) = 0$ if and only if $P \nabla f(d) = 0$.*

Proof. Suppose $P_d \nabla f_d(e) = 0$. Since $\phi_d(e) = \phi(d)$, from the application of Lemma 3.1 to $f_d(x)$ and $\psi_d(x)$, there exists $\alpha > 0$ such that $P_d \nabla \psi_d(\alpha e) = 0$. Applying Lemma 1.2, we have $P \nabla \psi(\alpha d) = 0$. Applying Lemma 3.1 again, we get $P \nabla f(d) = 0$. The converse follows in a similar fashion. ∎

Consider the projective transformation

$$T_d(x) = \left(\frac{n_1 D_1^{-1} x^{(1)}}{e^{(n_1)T} D_1^{-1} x^{(1)}}, \dots, \frac{n_m D_m^{-1} x^{(m)}}{e^{(n_m)T} D_m^{-1} x^{(m)}} \right)$$

which maps S one-to-one and onto itself. Its inverse is given by

$$T_d^{-1}(x) = \left(\frac{n_1 D_1 x^{(1)}}{e^{(n_1)T} D_1 x^{(1)}}, \dots, \frac{n_m D_m x^{(m)}}{e^{(n_m)T} D_m x^{(m)}} \right).$$

LEMMA 4.2. *Let d be a point in $W \cap \mathbb{R}^n_+$. For all $x \in W \cap \mathbb{R}^n_+$ we have*

$$\frac{f(T_d^{-1}(x))}{f(d)} = \frac{f_d(x)}{f_d(e)}.$$

Proof. From (3.1) and the identity $\pi(Dx) = \pi(d)\pi(x)$, we get

$$f(T_d^{-1}(x)) = f(Dx) = \frac{\phi(Dx)}{\pi(Dx)} = \frac{\phi_d(x)}{\pi(x)\pi(d)} = \frac{f_d(x)}{\pi(d)}.$$

The proof now follows by dividing the above by $f(d)$ and noting that $f_d(e) = \phi(d)$. ■

According to Lemma 4.2, to decrease the value of f at d it suffices to decrease that value of f_d at e. We now describe the algorithm. Assume that $\phi(d_0) > 0$ for some $d_0 \in W \cap \mathbb{R}^n_+$. Initialize $k = 0$.

Step 1. If $P_{d_k} \nabla f_{d_k}(e) = 0$, stop.

Step 2. Let $u_k = -P_{d_k} \nabla f_{d_k}(e)/\|P_{d_k} \nabla f_{d_k}(e)\|$ (steepest-descent direction of f_{d_k} at e). Let $\hat{\alpha}_k = \max\{\alpha : e + \alpha u_k \geqslant 0\}$. If $\phi_{d_k}(e + \alpha u_k)$ has a root α in $[0, \hat{\alpha}_k]$, stop.

Step 3. Let α_k be the smallest $\alpha \in (0, \hat{\alpha}_k)$ for which $\nabla f_{d_k}(e + \alpha_k u_k)^T u_k = 0$, i.e. the smallest stationary point of the function $g_k(\alpha) = f_{d_k}(e + \alpha u_k)$ in the interval $(0, \hat{\alpha}_k)$. Let $x_k = e + \alpha_k u_k$ and $d_{k+1} = T_{d_k}^{-1}(x_k)$. Set $k = k + 1$ and go to step 1.

THEOREM 4.1. *The steps of the algorithm are well defined. Moreover we have:*

(i) *If the algorithm stops while executing step* 1, *then* $P \nabla f(d_k) = 0$.

(ii) *If the algorithm stops while executing step* 2, *then* $V \neq \varnothing$.

(iii) *Suppose that the algorithm does not halt. Let $\hat{d}$ be an accumulation point of the sequence of d_k's. Then $\phi(\hat{d}) \geqslant 0$. If $\phi(\hat{d}) > 0$, then $\hat{d} \in W \cap S_+$ and $P \nabla f(\hat{d}) = 0$.*

Proof. Suppose $P_{d_k} \nabla f_{d_k}(e) \neq 0$; then u_k is well defined. We first prove that for any scalar α we have

$$e + \alpha u_k \in \hat{S} = \left\{x \in \mathbb{R}^n : e^{(n_j)^T} x^{(j)} = n_j, j = 1, \ldots, m\right\}. \quad (4.1)$$

The set $\hat{S}$ is obtained from S by relaxation of the nonnegativity constrains. Recalling that $P_{d_k} = \mathrm{diag}(P_{d_k}^{(1)}, \ldots, P_{d_k}^{(m)})$, the jth component of u_k is given by

$$u_k^{(j)} = -\frac{P_{d_k}^{(j)} \nabla_j f_{d_k}(e)}{\| P_{d_k} \nabla f_{d_k}(e) \|}.$$

To prove (4.1) it suffices to show $e^{(n_j)^T} u_k^{(j)} = 0$, for all $j = 1, \ldots, m$. Equivalently, since $P_{d_k}^{(j)} e^{(n_j)} = e^{(n_j)}$, it suffices to prove

$$e^{(n_j)^T} \nabla_j f_{d_k}(e) = 0 \qquad \forall j = 1, \ldots, m. \tag{4.2}$$

As in the derivation of (3.2) (essentially by substitution of f_{d_k} for f and e for x), it can be shown that

$$\nabla_j f_{d_k}(e) = \nabla_j \phi_{d_k}(e) - \sigma \phi_{d_k}(e) \lambda^{(j)} \qquad \forall j = 1, \ldots, m. \tag{4.3}$$

Multiplying both sides of (4.3) by $e^{(n_j)T}$, applying Property 1.2 to ϕ_{d_k}, and noting that $\sigma = K_j / \|\lambda^{(j)}\|_1$, we get the proof of (4.2).

From (4.1), the point $e + \alpha u_k$, whenever nonnegative, lies in S. Since S is compact, it follows that in step 2, $\hat{\alpha}_k$ is well defined and finite. Suppose that in step 2, ϕ_{d_k} has no zeros in $[0, \hat{\alpha}_k]$. Then, since $g_k'(0) < 0$, and $g_k(\alpha)$ approaches infinity as α approaches $\hat{\alpha}_k$, it follows that in the interval $(0, \hat{\alpha}_k)$, $g_k(\alpha)$ has at least one stationary point, and the least stationary point α_k is well defined. Also $x_k = e + \alpha_k u_k$ is in S. Thus, all the steps of the algorithm are well defined. Next we prove (i)–(iii).

(i): Suppose $P_{d_k} \nabla f_{d_k}(e) = 0$. Then $\phi(d_k) > 0$, and by Lemma 3.1, $P\, \nabla f(d_k) = 0$.

(ii): If $\phi_{d_k}(e + \alpha u_k) = 0$ for some $\alpha \in [0, \hat{\alpha}_k]$, then V_{d_k}, and hence V, is nonempty.

(iii): Suppose the algorithm does not halt. Since d_k is in $W \cap S_+$ for all k, we have $\hat{d} \in W \cap S$. Since $\phi(d_0) > 0$, we must have $\phi(d_k) = \phi_{d_k}(e) > 0$ for all k. Otherwise, the algorithm would have terminated while executing step 2. Thus, $\phi(\hat{d}) \geqslant 0$. Suppose $\phi(\hat{d}) > 0$. If $\hat{d} \notin W \cap S_+$, then $\pi(\hat{d}) = 0$. But this contradicts the monotonicity of the sequence of potential function values $f(d_k)$, which in turn is guaranteed by the fact that $f_{d_k}(x_k) < f_{d_k}(e)$ and by Lemma 4.2. Thus, $\phi(\hat{d}) > 0$ implies $\hat{d} \in W \cap S_+$. Furthermore, we will prove that in this case we also have

$$P_{\hat{d}}\, \nabla f_{\hat{d}}(e) = 0. \tag{4.4}$$

From Lemma 4.1, it then follows that $P\, \nabla f(\hat{d}) = 0$. The approach in proving (4.4) is analogous with the general approach in proving the convergence of the steepest descent using an appropriate step size; see e.g. McCormick [25]. However, our proof does take into account the specific properties of the projective algorithm.

Let N be an infinite subset of natural numbers such that for $k \in N$, the sequences d_k, x_k, u_k, and α_k converge to $\hat{d}$, $\hat{x}$, $\hat{u}$, and $\hat{\alpha}$, respectively. From step 3 of the algorithm we have

$$0 = \nabla f_{d_k}(x_k)^T u_k = \nabla f_{d_k}(e)^T P_{d_k} \nabla f_{d_k}(x_k) = \nabla f_{d_k}(e)^T P_{d_k} \nabla f_{d_k}(e + \alpha_k u_k). \tag{4.5}$$

If (4.4) is not satisfied, we must have $\hat{\alpha} > 0$. Otherwise, taking the limit in (4.5) as $k \in N$ approaches infinity, and using the continuity of $\nabla f_d(x)$ and P_d as functions of x or d, together with the fact that $P_d^2 = P_d$ and that $\hat{d}$ is in $W \cap S_+$, we get

$$\| P_{\hat{d}} \nabla f_{\hat{d}}(e) \|^2 = 0.$$

Thus, if (4.4) is not satisfied, we must have $\hat{\alpha} > 0$, and $-\nabla f_{\hat{d}}(e) P_{\hat{d}} \nabla f_{\hat{d}}(e) < 0$. Then, again from the continuity of $\nabla f_d(x)$ and P_d, there exists an infinite subset N' and N, and real numbers $\delta < 0$, and $\alpha^* > 0$, such that for $k \in N'$ and any $\tau \in (0, \alpha^*)$ we have

$$\nabla f_{d_k}(e + \tau u_k)^T u_k < \delta < 0, \qquad 0 < \alpha^* \leqslant \alpha_k. \tag{4.6}$$

From (4.6), the mean-value theorem, and the fact that in step 3 α_k is the least stationary point of $g_k(\alpha)$, for $k \in N'$ we have

$$\begin{aligned} f_{d_k}(e + \alpha_k u_k) - f_{d_k}(e) &\leqslant f_{d_k}(e + \alpha^* u_k) - f_{d_k}(e) \\ &= \alpha^* \nabla f_{d_k}(e + \tau_k u_k)^T u_k < \alpha^* \delta, \end{aligned} \tag{4.7}$$

where $\tau_k \in (0, \alpha^*)$. Since $0 < \phi(\hat{d}) = \lim_{k \in N' \to \infty} f_{d_k}(e)$, from (4.7) there exists an infinite subset N'' of N' such that for $k \in N''$ we have

$$0 < \frac{f_{d_k}(e + \alpha_k u_k)}{f_{d_k}(e)} \leqslant \frac{f_{d_k}(e + \alpha^* u_k)}{f_{d_k}(e)} \leqslant 1 + \frac{\alpha^* \delta}{f_{d_k}(e)} \leqslant \gamma < 1. \tag{4.8}$$

Now from (4.8), the monotonicity of $f(d_k)$, and Lemma 4.2 we have

$$0 < \frac{f(\hat{d})}{f(d_0)} = \prod_{k=0}^{\infty} \frac{f(d_{k+1})}{f(d_k)} = \prod_{k=0}^{\infty} \frac{f_{d_k}(e + \alpha_k u_k)}{f_{d_k}(e)}$$

$$\leqslant \prod_{k \in N''} \frac{f_{d_k}(e + \alpha_k u_k)}{f_{d_k}(e)} \leqslant \prod_{k \in N''} \gamma = 0,$$

a contradiction. Thus, (4.4) is satisfied. ■

From Lemma 3.1, if ψ has a unique stationary point in $W \cap \mathbb{R}_+^n$, then f must have a unique stationary point in $W \cap S_+$. Using this and Theorem 4.1, we conclude

THEOREM 4.2. *Assume ψ is convex over W. If $V = \varnothing$, then the sequence of d_k's in the projective algorithm is a convergent sequence.*

REMARK 4.1. For $m = 1$, $\phi(x) = c^T x$, and $\lambda = e$, the above algorithm reduces to a modified version of Karmarkar linear programming algorithm [17]. The modification being that a line search is carried out within each iteration. For $m = 2$, and ϕ a bilinear form (say) $y^T Qx$, where Q is an $m \times n$ positive matrix, the corresponding projective algorithm becomes a variant of the well-known RAS algorithm, which is the method of alternative row and column normalization; see e.g. [2], [4], [9], [32], [33], and [30] for a survey. The RAS algorithm, which is applicable to any m-dimensional nonnegative matrix, can also be considered as the coordinate descent algorithm applied to ψ. Equivalently, since Lemma 3.1 can also be stated in terms of partial gradients of ψ and f, the RAS can also be considered as the coordinate descent as applied to f. Thus, while the projective algorithm of this paper uses the total gradient in its descent directions, the RAS uses partial gradients. Hence, one would expect a better reduction of the potential function in each iteration of the projective algorithm than that obtained from each complete cycle of iterations of the RAS algorithm (i.e. descent with respect to all the variables). This has been witnessed in some preliminary computational results with nonnegative matrices. However, for nonnegative matrix scaling the projective algorithm requires a line search (in computing the step length α_k), while the RAS does not. Nevertheless, the cost of line search is nominal and is dominated by the cost of other operations. For example, for the problem of scaling of an $n \times n$ nonnegative matrix, the cost of each complete cycle of iterations of the RAS is $O(n^2)$, and the cost of each iteration of the projective algorithm for the same problem is again $O(n^2)$, plus the cost of a linear search. The analogy between the RAS and the projective algorithm for nonnegative matrices suggests that for a general multihomogeneous function, an RAS version of the projective algorithm can be stated, i.e. one that uses projected partial gradients as descent directions.

Many important and interesting research problems can be stated with respect to both the theorem of the alternative and the projective algorithm whose convergence properties have been based only on continuous differentiability and multihomogeneity of ϕ. One such problem is the analysis of theoretical and practical rate of convergence of the algorithm. Clearly, such an analysis only makes sense when ϕ is restricted to lie within specific classes of multihomogeneous functions. A particular case of interest is to obtain such an analysis for nonnegative matrices, as well as its comparison with the RAS algorithm, for which such theoretical results have already been obtained in [16]. A second problem of interest is the characterization of the cases of ϕ for which the exclusivity of conditions (i) and (ii) in Theorem 1.1 holds true. For such cases the projective algorithm can decisively determine which of the two conditions holds. For instance for the case of $m = 1$, in [12] it is shown that the exclusivity of (i) and (ii) holds true for homogeneous functions which are copositive plus, a class of homogeneous functions which contain the convex ones. However, even for symmetric matrices, this exclusivity holds for a class of matrices which is strictly and nontrivially larger than the class of copositive plus matrices (see [12]). Nevertheless, also in the case of symmetric matrices a total characterization of the above exclusivity is not available.

I would like to thank Professor Louis Billera for stimulating discussions regarding a homogeneous version of this paper given in [12], Professors Omer Egecioglu and Marvin Marcus for their communications regarding some literature on nonnegative matrix scaling, and the editor Professor Hans Schneider and the referees for bringing to my attention other literature on nonnegative matrix scaling and for suggestions that resulted in the improvement of the presentation.

REFERENCES

1 M. Bacharach, *Biproportional Matrices and Input-Output Change*, Cambridge U.P., (1970).

2 A. Bachem and B. Korte, On the RAS algorithm. *Computing* 23:189–198 (1979).

3 R. B. Bapat and T. E. S. Raghavan, An extension of a theorem of Dorrah and Ratcliff in loglinear models and its application to scaling multidimensional matrices, *Linear Algebra Appl.* 114/115:705–715 (1989).
4 L. M. Bregman, Proof of the convergence of Sheleikhovskii's method for a problem with transportation constraints, *USSR Comput. Math. and Math. Phys.* 7:191–204 (1967).
5 R. A. Brualdi, S. V. Parter, and H. Schneider, The diagonal equivalence of a nonnegative matrix to a stochastic matrix, *J. Math. Ann. Appl.* 16:31–50 (1966).
6 R. A. Brualdi, Convex sets of non-negative matrices, *Canad. J. Math.* 20:144–157 (1968).
7 G. B. Dantzig, *Linear Programming and Extensions*, Princeton, U.P., Princeton, N.J. (1963).
8 D. Z. Djoković, Note on nonnegative matrices, *Proc. Amer. Math. Soc.* 25:80–82 (1970).
9 J. Franklin and J. Lorenz, On the scaling of multidimensional matrices, *Linear Algebra Appl.* 114/115:717–735 (1989).
10 B. Kalantari, Karmarkar's algorithm with improved steps, *Math. Programming* 46:73–78 (1990).
11 B. Kalantari, Canonical problems for quadratic programming and projective methods for their solution (presented at AMS conference Mathematical problems arising from linear programming, 1988), *Comtemp. Math.* 114:243–263 (1990).
12 B. Kalantari, Derivation of a generalized and strengthened Gordan theorem from generalized Karmarkar potential and logarithmic barrier functions, Technical Report LCSR-TR-121, Dept. of Computer Science, Rutgers Univ., New Brunswick, N.J., 1989.
13 B. Kalantari, Generalization of Karmarkar's algorithm to convex homogeneous functions, *Oper. Res. Lett.* 11:93–98 (1992).
14 B. Kalantari, A simple polynomial time algorithm for a convex hull problem equivalent to linear programming, Combinatorics Advances, Kluwer Academic Pub., 1995, to appear.
15 B. Kalantari and L. Khachiyan, On the complexity of nonnegative matrix scaling, Technical Report LCSR-TR-156, Dept. of Computer Science, Rutgers Univ., New Brunswick, N.J., 1990.
16 B. Kalantari and L. Khachiyan, On the rate of convergence of deterministic and randomized RAS matrix scaling method, *Oper. Res. Lett.* 14:237–244.
17 N. Karmarkar, A new polynomial time algorithm for linear programming, *Combinatorica* 4:373–395 (1984).
18 L. Khachiyan, A polynomial algorithm in linear programming, *Soviet Math. Dokl.* 20:191–196 (1979).
19 L. Khachiyan, and B. Kalantari, Diagonal matrix scaling and linear programming, *SIAM J. Optim.* 2:668–672 (1992).
20 L. Khachiyan, Matrix scaling in NP-hard, Technical Report LCSR-TR-287, Dept. of Computer Science, Rutgers Univ., New Brunswick, N.J., 1991.
21 D. London, On matrices with a doubly stochastic pattern, *J. Math. Anal. Appl.* 34:648–652 (1971).

22 M. Marcus and M. Newman, The permanent of a symmetric matrix, *Notices Amer. Math. Soc.* 8:595 (1961).
23 A. W. Marshall and I. Olkin, Scaling of matrices to achieve specified row and column sums, *Numer. Math.* 12:83–90 (1968).
24 J. Maxfield and H. Minc, A doubly stochastic matrix equivalent to a given matrix, *Notices Amer. Math. Soc.* 9:309 (1962).
25 G. P. McCormick, *Nonlinear Programming Theory, Algorithms and Applications*, Wiley, New York, 1983.
26 M. V. Menon, Reduction of matrix with positive elements to a doubly stochastic matrix, *Proc. Amer. Math. Soc.* 18:244–247 (1967).
27 H. Minc, *Nonnegative Matrices*. Wiley, New York, 1988.
28 L. Mirsky and H. Perfect, The distribution of positive elements in doubly stochastic matrices, *J. London Math. Soc.* 49:689–698 (1965).
29 U. G. Rothblum and H. Schneider, Scaling of matrices which have prescribed row sums and column sums via optimization, *Linear Algebra Appl.* 114/115:737–764 (1989).
30 M. H. Schneider and S. A. Zenios, A comparative study of algorithms for matrix balancing, *Oper. Res.* 38:439–455 (1989).
31 M. H. Schneider, Matrix scaling, entropy minimization, and conjugate duality. I. Positivity conditions, *Linear Algebra Appl.* 114/115:785–813 (1989).
32 R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.* 35:876–879 (1964).
33 R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.* 21:343–348 (1967).

Received 14 April 1993; final manuscript accepted 11 July 1994